

Parametric coding of an audio or speech signal

The invention relates to a parametric encoder and method for encoding an audio or speech signal into sinusoidal code data according to the preambles of claims 1 and 6, respectively.

The invention further relates to a parametric decoder and method for re-
5 constructing an approximation of said audio or speech signal from said sinusoidal code data according to the preambles of claims 11 and 12, respectively.

Audio and speech signals are preferably encoded before being transmitted via a channel or stored on a storage medium in order to compress the data of said signals. Audio or speech signals are substantially represented by sinusoidal code data and consequently
10 specific encoders are known in the art specialised for the encoding of these signals. Such a parametric encoder is e.g. known from E.B. George and M.J.T. Smith, "A new speech coding model based on a least-squares sinusoidal representation". In Proc. 1987 Int. Conf. Acoust. Speech Signal Process. (ICASSP87), pages 1641-1644, Dallas TX, 6-9 April 1987. IEEE, Picataway, NJ. The parametric encoder described there is illustrated in Fig. 5. According to
15 Fig. 5 the parametric encoder 500 comprises a segmentation unit 510 for segmenting a received audio or speech signal s into at least one finite segment $x(n)$.

Said segment $x(n)$ is input to a calculation unit 520. Said calculation unit 520 calculates sinusoidal code data in the form of phase and amplitude data of a given extension \hat{x} from the segment $x(n)$ such that the extension \hat{x} approximates the segment $x(n)$ as good
20 as possible for a given criterion, e.g. minimum of weighted squared error. For the cited parametric encoder the extension is given by

$$\hat{x}(n) = \sum_{l=1}^L A'_l(n) \cos(\Phi'_l(n)) \quad (1)$$

with

$$25 \quad A'_l(n) = \sum_{j=0}^{J-1} \alpha'_j n^j \quad (2)$$

$$\Phi'(n) = \sum_{k=0}^{K-1} \phi'_k n^k \quad (3)$$

with α'_j and ϕ'_k are polynomial coefficients of the amplitude parameter A^i and of the phase parameter Φ' .

5 The calculation unit 520 comprises a frequency estimation unit 522 for calculation the phase coefficients ϕ'_k from the received segment $x(n)$ for example, for $k = 1$ (thus ϕ'_1), by picking frequencies in the frequency spectrum of said segment $x(n)$. These phase coefficients ϕ'_k represent the phase part of said sinusoidal code data are on one hand output to a multiplexer 530 and are on the other hand input into a pattern generation unit 524.

10 Said pattern generation unit serves for calculating the phase parameter $\Phi'(n)$ according to equation (3).

The pattern generation unit 524 further generates a plurality of JxL components p_{ij} of the extension $\hat{x}(n)$ according to

$$15 \quad p_{ij}(n) = n^j \cos(\Phi^i(n)), \text{ with } i = 1-L, j = 0-(J-1)$$

The plurality of JxL components p_{ij} is input to an amplitude estimation unit 526 which determines the optimal amplitude data α'_j from said received components as well as from the received segment $x(n)$ output from the segmentation unit 510.

20 The phase coefficients ϕ'_k and the amplitudes α'_j form the sinusoidal code data which represents the extension $\hat{x}(n)$ as an approximation of the segment $x(n)$. These sinusoidal code data are multiplexed by the multiplexer 530 in order to form a data stream which may be stored on a recording medium or transmitted via a channel.

The extension $\hat{x}(n)$ as described by equation 1 and as known from the

25 described parametric encoder 500 provides a proper approximation for an individual segments $x(n)$ of the audio or speech signal. However, the calculation of the sinusoidal code data is rather complicated.

Starting from that prior art it is an object of the invention to improve a known parametric encoder and method for encoding an audio or speech signal into sinusoidal code

30 data and to improve a known parametric decoder and method for re-constructing an

approximation of said audio or speech signal from said sinusoidal code data after transmission or restoration such that the calculation of said sinusoidal code data can be carried out in a simpler and cheaper way.

This object is solved by the subject matter of claim 1. More specifically, this

- 5 object is solved by adapting the calculation unit to calculate the sinusoidal code data θ'_k, d'_j and e'_j for the following extension \hat{x} :

$$\hat{x}(n) = \sum_{i=1}^L \sum_{j=0}^{J-1} [d'_j f_j(n) \cos(\Theta'(n)) + e'_j f_j(n) \sin(\Theta'(n))]$$

with

10 $\Theta'(n) = \sum_{k=1}^K \theta'_k n^k$

wherein:

- i : represents a component of the extension $\hat{x}(n)$;
- j,k : represent parameters;
- n : represents a discrete time parameter;
- 15 θ'_k : represents the phase coefficient value as one of said sinusoidal code data
- f_j : represents the jth instance out of the set of J linearly independent functions;
- Θ^i : is a phase; and
- 20 d'_j, e'_j : represent the linearly involved amplitude values of the components representing the amplitude parts of said sinusoidal code data.

- Advantageously, the optimisation problem occurring when trying to define the
- 25 sinusoidal data such that the claimed extension \hat{x} accurately describes a specific segment $x(n)$ is easy to solve. The easy calculation results from the fact that except the phase coefficients θ'_k the amplitude data d'_j and e'_j are linearly involved within the claimed extension \hat{x} . Note that there does not appear a zeroth order phase coefficient in Θ' , whereas such component exists in Φ' in the form of ϕ'_0 .

Further, advantageously the claimed extension \hat{x} provides more degrees of freedom for defining the sinusoidal code data with the result, that the claimed extension \hat{x} is broader than the extensions known in the art and provides a more accurate approximation of an individual segment $x(n)$.

According to a first embodiment of the invention the linearly independent function $f_j(n)$ is set to $f_j(n) = n^j$. In that way the claimed extension \hat{x} is restricted to a polynomial extension.

Further advantageous embodiments of the claimed parametric encoder and in particular of the claimed calculation unit are subject matter of the dependent encoder claims.

The above identified object is further solved by a method for encoding an audio or speech signal as claimed in claim 6. The advantages and embodiments of the said method correspond to the advantages and embodiments as explained above for the parametric encoder.

The above identified object is further solved by a parametric decoder for re-constructing an approximation \hat{x} of an audio or speech signal from transmitted or restored code data according to claim 11. More specifically, the object is solved by adapting a known synthesiser to re-construct said segments \hat{x} from said sinusoidal code data ϕ_k^i , d_j^i and e_j^i according to the following formula:

$$\hat{x}(n) = \sum_{i=1}^L \sum_{j=0}^{J-1} [d_j^i f_j(n) \cos(\Theta^i(n)) + e_j^i f_j(n) \sin(\Theta^i(n))]$$

with

$$\Theta^i(n) = \sum_{k=1}^K \theta_k^i n^k$$

wherein:

- | | | |
|---------------|---|---|
| i | : | represents a component of the extension $\hat{x}(n)$; |
| 25 j,k | : | represent parameters; |
| n | : | represents a discrete time parameter; |
| f_j | : | represents the jth instance out of the set of J linearly independent functions; |
| θ_k^i | : | represents the phase coefficient as one of said sinusoidal data |
| 30 Θ^i | : | is a phase parameter; and |

d'_j, e'_j : represent the linearly involved amplitude values of the components representing parts of said sinusoidal data.

Advantageously, the calculation of the claimed extension \hat{x} is easier than the calculation of the extensions known in the art. This is due to the linear involvement of the amplitude data d'_j and e'_j within said extension and the omission of the zeroth-order phase coefficient.

Due to the easy calculation of the extension \hat{x} the reconstruction of the original audio or speech signal s in the form of its approximation \hat{x} can be realised cheaper and quicker.

The above identified object is further solved by the decoding method as claimed by claim 12. The advantages of said method correspond to the advantages mentioned above by referring to the parametric decoder.

Five figures are accompanying the description, wherein

Fig. 1 shows a first embodiment of the parametric encoder according to the invention;

Fig. 2 shows a second embodiment of the parametric encoder according to the invention;

Fig. 3 shows a flow chart illustrating the operation of the second embodiment of the parametric encoder according to the invention;

Fig. 4 shows a parametric decoder according to an embodiment of the invention; and

Fig. 5 shows a parametric encoder as known in the art.

Before describing the preferred embodiments of the invention some basic explanations about the subject matter of the invention are given.

The invention proposes an extension $\hat{x}(n)$ for approximating a segment $x(n)$ of a sinusoidal audio or speech signal s . Said extension $\hat{x}(n)$ is represented by phase and amplitude data, hereinafter also referred to as sinusoidal code data. The sinusoidal code data is defined such that the extension $\hat{x}(n)$ approximates the segment $x(n)$ of the audio or speech signal as good as possible for a given criterion, e.g. minimisation of the squared weighted error. Expressed in other words, the sinusoidal code data has to be defined by

solving an optimisation problem. After the sinusoidal code data has been defined for optimally approximating a particular segment $x(n)$ it might be stored on a storage medium or transmitted via a channel as code data representing said segment $x(n)$ and thus also representing said audio or speech signal s . Preferably, before being stored or transmitted the sinusoidal code data might be encoded and/or cleaned in the way that irrelevant or redundant data is removed from it.

The generation of said sinusoidal code data according to a first embodiment is now explained by referring to Fig. 1.

Fig. 1 shows a first preferred embodiment of a parametric encoder 100 for generating said sinusoidal code data representing an input audio or speech signal s . The received signal s is input to a segmentation unit 110 for segmenting said signal s into at least one segment $x(n)$. Said segment $x(n)$ is input into a calculation unit 120 for generating said sinusoidal code data such that the extension \hat{x} with

$$\hat{x}(n) = \sum_{i=1}^L \sum_{j=0}^{J-1} [d'_j f_j(n) \cos(\Theta^i(n)) + e'_j f_j(n) \sin(\Theta^i(n))] \quad (4)$$

with

$$\Theta^i(n) = \sum_{k=1}^K \theta_k^i n^k \quad (5)$$

and wherein:

- i, j, k : represent parameters;
- n : represents a discrete time parameter;
- θ_k^i : represents the phase coefficient as one of said sinusoidal data
- f_j : represents the j th instance out of the set of J linearly independent functions;
- Θ^i : is a phase; and
- d'_j, e'_j : represent the linearly involved amplitude values of the components representing parts of said sinusoidal data

approximates the segment $x(n)$ input to said calculation unit 120 as good as possible for a given criterion, e.g. minimisation of weighted squared error. The sinusoidal code data to be determined by said calculation unit 120 is the phase θ'_k and the amplitude data d'_j and e'_j .

The term C_i of equation (4) with

$$C_i = \sum_{j=0}^{J-1} [d'_j f_j(n) \cos(\Theta'(n)) + e'_j f_j(n) \sin(\Theta'(n))] \quad (6)$$

is hereinafter referred to as the i 'th component of the extension \hat{x} with $i = 1-L$.

The calculation unit 120 comprises a frequency estimation unit 122 for determining a plurality of $L \times K$ phase coefficients θ'_k with $k = 1-K$ for all components C_i with $i = 1-L$ of the extension $\hat{x}(n)$ according to formula (5) representing the individually received segment $x(n)$. Said plurality of $L \times K$ frequencies θ'_k is input to a pattern generating unit 124 for calculating a plurality of L frequency parameters $\Theta'(n)$ with $i = 1-L$ according to formula (5). Said pattern generating unit 124 is further adapted for generating a plurality of $J \times L$ pairs of patterns p_y^1, p_y^2 , for the components C_i with $i = 1-L$ according to:

$$p_{ij}^1 = f_j(n) \cos(\Theta'(n)); \text{ and}$$

$$p_{ij}^2 = f_j(n) \sin(\Theta'(n))$$

for $i = 1-L$ and $j = 0-(J-1)$.

Said plurality of pairs of patterns p_y^1, p_y^2 is - together with the segment $x(n)$ - input to an amplitude estimation unit 126 for determining a plurality of $J \times L$ amplitude data d'_j for all received patterns p_y^1 and a plurality of $J \times L$ amplitude data e'_j for all the received patterns p_y^2 of all components C_i of the extension $\hat{x}(n)$.

The calculation unit 120 and in particular the frequency estimation unit 122 and the amplitude estimation unit 126 are adapted such that the sinusoidal data comprising the phase data θ'_k and the amplitude data d'_j and e'_j is determined and optimised such that the criterion "minimisation of weighted squared error E between the segment $x(n)$ and the extension $\hat{x}(n)$ " is (approximately) fulfilled.

The parametric encoder 100 may further comprise a multiplexer 130 for transforming the plurality of $L \times K$ phase coefficients θ_k^i as output by said frequency estimation unit 122 and said plurality of $J \times L$ amplitude data d_j^i and e_j^i as output by said amplitude estimation unit 126 into a data stream to be stored on a storage medium or to be transmitted via a channel.

Fig. 2 shows a second embodiment of the parametric encoder 100'. Like the parametric encoder 100 the parametric encoder 100' also serves for generating said sinusoidal code data from the input audio or speech signal s . The operation of its segmentation unit 110' corresponds to the operation of the segmentation unit 110.

Consequently, the segmentation unit 110' generates segments $x(n)$ of the received signal s at its output. Said segments $x(n)$ are input to a calculation unit 120'. In difference to the first embodiment of the calculation unit 120 the calculation unit 120' does not calculate the plurality of sinusoidal code data simultaneously for all components of a segment $\hat{x}(n)$ but generates this sinusoidal code data sequentially for each component C_i with $i = 1-L$ of the extension \hat{x} . This way of calculation is generally known in the art as analysis-by-synthesis or as matching pursuit algorithm. However, in the prior art an application of said method is only known for extensions different from the claimed extension \hat{x} according to formula (4).

In the following the operation of said second embodiment of the calculation unit 120' is explained by referring to Figs. 2 and 3. More specifically, the calculation of the sinusoidal code data of the extension \hat{x} according to equation (4) is described such that the weighted squared error between a segment output by the segmentation unit 110' and its extension \hat{x} according to equation (4) is (approximately) minimised.

In a first cycle $i = 1$ the sinusoidal code data of a first component C_i with $i = 1$ of the extension \hat{x} are calculated (method step a) in Fig. 3).

For achieving this, the output of segmentation unit 110' $x(n)$ is set to: $\varepsilon_{i-1} = x(n)$ (see method step b)).

In said first cycle, said output of the segmentation unit 110' is input to a frequency estimation unit 122' for determining a plurality of K phase coefficients θ_k^i with $k = 1-K$ from the input value ε_{i-1} (see method step c)). Said phase coefficients θ_k^i represent the phases of the searched sinusoidal code data and are thus output from the calculation unit. Moreover, said phase coefficients θ_k^i are input to a pattern generating unit 124' for calculating the phase Θ^i with $i = 1$ for the first component C_1 according to equation (5) (see

method step d)). Said pattern generating unit 124' further serves for generating a plurality of $2 \times J$ patterns with $j = 0-(J-1)$ for the component C_i with:

$$p_{ij}^1 = f_j(n) \cos(\Theta'(n)); \text{ and}$$

$$5 \quad p_{ij}^2 = f_j(n) \sin(\Theta'(n))$$

for $i = 1$ (see method step e)). These generated patterns p_{ij}^1, p_{ij}^2 are - together with the parameter ε_{i-1} - input to an amplitude estimation unit 126'. Said amplitude estimation unit 126' serves for determining a plurality of J amplitudes d_j' for said patterns p_{ij}^1 and of J amplitudes e_j' for said patterns p_{ij}^2 for the component C_i with $i = 1$ from the received input data (see method step f)). Said calculated amplitudes d_j' and e_j' form the amplitude part of the sinusoidal data representing the extension \hat{x} of the segment $x(n)$ and are thus output from that calculation unit 120' in order to be - together with said phase data θ_k' merged into a data stream representing said first component C_i with $i = 1$. Moreover, said amplitude data d_j' and e_j' are - together with their respective patterns p_{ij}^1 and p_{ij}^2 input into a synthesiser 128' for calculating the component C_i with $i = 1$ according to

$$C_i = \sum_{j=0}^{J-1} [d_j' f_j(n) \cos(\Theta'(n)) + e_j' f_j(n) \sin(\Theta'(n))]$$

(see method step g)).

Said component C_i is input into a subtracting unit 129' for being subtracted from the value ε_{i-1} being input to said frequency estimation unit 122'. The difference occurring at the output of said subtracting unit 129' is referred to as ε_i with $i = 1$ (see method step h)).

Now the first cycle for calculating the first component C_1 and its sinusoidal code data θ_k', d_j' and e_j' for the extension \hat{x} has been finished. Subsequently, the parameter i is compared with the total number L of components C_i of the segment \hat{x} (see method step i)). If $i < L$ method steps c) to i) are repeated for $i = i+1$. In these cases the output from the segmentation unit 110' for $i \geq 1$ is disconnected from the input of the frequency estimation unit 122'; instead, the input of said frequency estimation unit 122' is connected to the output of said subtracting unit 129' for receiving the differences ε_i . However, if $i \geq L$ the

sinusoidal code data of all L components of the extension \hat{x} have been calculated and thus the calculation process carried out by the calculation unit 120' has been finished for a particular segment \hat{x} . Subsequently, the whole procedure may be repeated for a subsequent segment of the input audio or speech signal.

5 Fig. 4 shows a parametric decoder 400 for reconstructing an approximation \hat{s} of an audio or speech signal s from received input data. These received input data correspond to data of a data stream after being transmitted or restored from a storage medium.

The parametric decoder 400 comprises a selecting unit 420 for selecting sinusoidal code data θ'_k , d'_j and e'_j representing segments \hat{x} of the approximation \hat{s} of the
10 audio and/or speech signal s from said received input data. The parametric decoder 400 further comprises a synthesiser 440 for reconstructing said segments \hat{x} from said received sinusoidal code data and a joining unit 460 for re-constructing the approximation \hat{s} by linking the re-constructed segment \hat{x} .

It should be noted that the above-mentioned embodiments illustrate rather than
15 limit the invention, and that those skilled in the art will be able to design many alternative embodiments without departing from the scope of the appended claims. In the claims, any reference signs placed between parentheses shall not be construed as limiting the claim. The word 'comprising' does not exclude the presence of other elements or steps than those listed in a claim. The invention can be implemented by means of hardware comprising several
20 distinct elements, and by means of a suitably programmed computer. In a device claim enumerating several means, several of these means can be embodied by one and the same item of hardware. The mere fact that certain measures are recited in mutually different dependent claims does not indicate that a combination of these measures cannot be used to advantage.